Language Learning & Technology
http://llt.msu.edu/vol12num3/rimrottheift/
October 2008, Volume 12, Number 3
pp. 73-92

EVALUATING AUTOMATIC DETECTION OF MISSPELLINGS IN GERMAN

Anne Rimrott and **Trude Heift**

Simon Fraser University

This study investigates the performance of a spell checker designed for native writers on misspellings made by second language (L2) learners. It addresses two research questions: 1) What is the correction rate of a generic spell checker for L2 misspellings? 2) What factors influence the correction rate of a generic spell checker for L2 misspellings? To explore these questions, the study considers a corpus of 1,027 unique misspellings from 48 Anglophone learners of German and classifies these along three error taxonomies: linguistic competence (competence versus performance misspellings), linguistic subsystem (lexical, morphological or phonological misspellings), and target modification (single-edit misspellings (edit distance = one) versus multiple-edit misspellings (edit distance > 1)). The study then evaluates the performance of the *Microsoft Word*® spell checker on these misspellings. Results indicate that only 62% of the L2 misspellings are corrected and that the spell checker, independent of other factors, generally cannot correct multiple-edit misspellings although it is quite successful in correcting single-edit errors. In contrast to most misspellings by native writers, many L2 misspellings are multiple-edit errors and are thus not corrected by a spell checker designed for native writers. The study concludes with computational and pedagogical suggestions to enhance spell checking in CALL.

INTRODUCTION

Along with speaking, reading, and listening, writing is one of the four essential skills of learning a foreign language (L2). Learners of all proficiency levels are confronted with the daunting task of writing in the L2 and they use what tools they can find to assist them in the process. For this reason, spell checkers have become very popular proofing tools in the L2 classroom.

To check their writing, L2 learners frequently use *generic* spell checkers, that is, spell checkers designed for native (L1), as opposed to nonnative, writers. Generic spell checkers, such as the spell checker included in the *Microsoft Word*® word processing package, are readily available and enjoy wide distribution. In contrast, spell checkers specifically designed for nonnative writers are rare, not well-integrated into widely-used word processing software, have limited distribution, and often represent additional purchasing costs.

Despite their popularity, the effectiveness of generic spell checkers in treating nonnative misspellings has not been empirically evaluated with the exception of a few more recent studies (e.g., Hovermale, 2008; Rimrott & Heift, 2005). Yet, in Computer-Assisted Language Learning (CALL), empirical studies involving a corpus of authentic L2 misspellings are needed for at least two reasons: first, to evaluate existing spell checkers, and, second, to inform the design of new L2 spell checkers by exposing misspellings that should be targeted in nonnative writing (Cowan, Choi & Kim, 2003; Dagneaux, Denness, & Granger, 1998; Granger & Meunier, 1994; Ndiaye & Vandeventer Faltin, 2003). Moreover, empirical studies can also reveal learner strategies for effective use of spell checkers (Heift & Rimrott, 2008).

The goal of the present study is to address the lack of empirical research on L2 spell checking in CALL by evaluating the effectiveness of a generic spell checker on a corpus of nonnative misspellings. Specifically, our study examines the performance of the generic spell checker in *Microsoft Word*® on

1027 misspellings produced by Anglophone learners of German. The following research questions are addressed:

- What is the correction rate of a generic spell checker for L2 misspellings?
- What factors influence the correction rate of a generic spell checker for L2 misspellings?

LITERATURE REVIEW

Effectiveness of Spell Checkers for Native Writers

Generic spell checkers, like the one in *Microsoft Word®,* target L1 writers and assume that misspellings mainly involve accidental mistypings, which are fairly predictable minimal deviations from the correct spellings (Dagneaux et al., 1998; Helfrich & Music, 2000; Pollock & Zamora, 1984). Specifically, the algorithms of generic spell checkers are largely based on the empirical finding that the vast majority of L1 misspellings involves an edit distance of one. Edit distance is defined as the number of additions, omissions, substitutions or transpositions needed to convert a misspelling into its target word. Accordingly, most L1 misspellings contain only a *single* error of omission (e.g., **spel/spell*), addition (**sspell*), substitution (**soell*), or transposition (**sepll*) (Damerau, 1964; Pollock & Zamora, 1984). Apart from problems with proper nouns, rare words, and real-word errors (e.g., **their/there*), generic spell checkers successfully handle the majority of misspellings made by typical native speakers. Kukich (1992) notes that "most researchers report accuracy levels above 90% when the first three guesses [in a spell checker's list of suggested corrections] are considered" (p. 412).[1]

When it comes to L2 misspellings, however, the spell checker's success rate is potentially quite different because L2 errors differ substantially from typical L1 errors. To illustrate this, the following subsection provides a definition for misspellings by also considering computational aspects of system flow when spell checking a text.

Challenges in Defining Misspellings

Word processors generally have a task division between the spell checker and the grammar checker. The spell checker scans a text for *all* unknown words, that is, all words that are not contained in the spell checker's dictionary. Once these words have been identified and resolved (e.g., the user modifies the original input or tells the system that the word is spelled correctly), the grammar checker examines the text for grammatical errors such as subject-verb agreement.[2]

In L2 writing, in addition to accidental mistypings, learners produce words unknown to the word processor due to misconceptions of the L2. For example, a learner might inflect a verb incorrectly, which results in a nonexistent word (e.g., **goed/went*). Because of the task division in word processors, studies with a computational component generally classify errors such as **goed/went* as *morphology-triggered misspellings* (Kese, Dudda, Heyer, & Kugler, 1992; L'haire, 2007; Ndiaye & Vandeventer Faltin, 2003; Rimrott & Heift, 2005) rather than grammar (morphological) errors. This tradition is also followed here. However, the fact that a spell checker handles these morphological errors does not imply that the suggestions for error repair should focus on spelling. On the contrary, learner feedback for a morphology-triggered misspelling ideally explains the morphological nature of the error without making reference to spelling. From a computational point of view, this, however, requires that the error can be identified as such by the computer program, which is generally not the case with a spell checker designed for native speakers.[3] In sum, a misspelling is a word that is not contained in the spell checker's dictionary. From a computational point of view, this most commonly implies that the word also does not exist in the target language given that the spell checker's dictionary should contain all of its legitimate words. Table 1 provides examples of misspellings in both German and English.

Table 1. Examples of Misspellings

Description	English example	German example
Wrongly inflected forms resulting in nonexistent words	**goed/went*	**gebst/gibst* 'give'
Incorrect use of lower case resulting in nonexisting words	**canada/Canada*	**nacht/Nacht* 'night'
Other nonexistent words (e.g., accidental mistypings)	**pn/pen*	**Schwrster/Schwester* 'sister'

Spell checking for Nonnative Writers

Spell checking text produced by language learners is considerably different from checking native speaker texts, not least because of the possibility of many misspellings triggered by grammatical misconceptions of the L2 (e.g., **goed*/*went*). However, only a few studies discuss the effectiveness of spell checkers in handling nonnative misspellings. Both Burston (1998) and Holmes and de Moras (1997) investigate the effectiveness of a French grammar and spell checker (*Antidote 98* and *Le Correcteur 101*, respectively) when tested against essays by English-L1 university students. Burston found that while *Antidote* dealt with most misspellings effectively, it misidentified "some fairly obvious spelling errors" (p. 209). Holmes and de Moras concluded that *Le Correcteur*'s spell checking "usefulness would be extended if it were taught to anticipate some typical Anglophone errors" (p. 104).

In a pilot study, Hovermale (2008) investigated 101 nonword spelling errors by Japanese learners of English, including morphology-triggered misspellings such as **holded/held*. The correction rate for the 101 misspellings was 79% and 81%, respectively, when tested using the *Microsoft Word® 2003* and *ASPELL 0.60* (http://aspell.net/) spell checkers.

In a small-scale study, Rimrott and Heift (2005) analyzed 374 spelling errors by 34 learners of German. Their study found that, unlike the fairly high success rate in targeting L1 misspellings, the generic spell checker they evaluated only corrected 52% of the nonnative misspellings. However, the spell checker was much more successful in treating misspellings that deviated less substantially from the target words, that is, errors that resembled those made by typical native speakers.

While not conducting an empirical analysis, Kese et al. (1992), nevertheless, note shortcomings of spell checkers in the L2 context:

> Many more errors could be detected by a spelling corrector if it possessed at least some rudimentary linguistic knowledge. … when confronted with a regular though false form of [an irregular word like "mouse"] (e.g. with … "mouses"), … a [standard] system normally fails to propose the corresponding irregular form (… "mice") as a correction alternative. (p. 126)

The studies above point to the limitations of generic spell checkers when it comes to correcting nonnative misspellings and, in response, several L2 spell checkers have been developed. Most of these programs specifically target certain errors classes, such as phonology-triggered or morphology-triggered misspellings. They may also anticipate errors by incorporating lists of commonly misspelled words in the target language (e.g., *Antidote*, a spell checker for learners of French). In addition, a CALL program may provide extra tools to address shortcomings commonly associated with generic spell checkers.

L2 spell checkers targeting phonological misspellings generally obtain a phonological representation of the misspelling and then retrieve from the dictionary correction alternatives with the same phonological representation or close approximates. *SPELLER* is a program that mainly targets phonologicallymotivated misspellings by Dutch learners of English (de Haan & Oppenhuizen, 1994). *CorText* (reviewed by Mydlarski, 1999), a grammar and spell checker for French learners of English, uses phonetic approximation to detect phonological misspellings.

There are also a number of spell checkers that target both phonological and morphological misspellings. An example of this is *FipsOrtho* (L'haire, 2007), a spell checker for learners of French that developed out of *FipsCor* (Ndiaye & Vandeventer Faltin, 2003). Regarding morphological misspellings, the program, for instance, can correct the incorrect plural regularization in **animals/animaux* 'animals'. Nadasdi and Sinclair's *Spellcheckplus* (http://spellcheckplus.com/) and *BonPatron* (http://bonpatron.com/) are online L2 English and L2 French spell and grammar checkers, respectively. In addition to correcting typographical and some phonological misspellings, the two programs are able to correct morphologically-triggered errors. SCALE (Spelling Correction Adapted for Learners of English) is an L2 spell checker currently under development that addresses phonological confusion and morphological overregularization (e.g., **feeled/felt*) by Japanese learners of English (Hovermale, 2008).

Het Spelraam (Bos, 1994), a tutoring system for the conjugation and spelling of Dutch verbs as a spelling aid for children or L2 learners, mainly addresses morphological misspellings. Some additional spell checkers target L2 French learners, for example, the educational version of *Sans-Faute* and *Le Correcteur Didactique* (both reviewed by Murphy-Judy, 2003).

Finally, there are L2 spell checkers that, instead of targeting certain error classes, provide learners with additional tools to overcome the limitations of generic spell checkers. For instance, *the Penguin* (Fallman, 2002) is a descriptive grammar and spell checker that uses the Internet as a reference database. If a learner is unsure of the spelling of a particular word, the number of hits for alternative spellings, as retrieved by a search engine, can be compared to determine the correct spelling (i.e., the alternative with the most hits is likely to be correct).

Despite these efforts, L2 spell checkers are rare and, for this reason, L2 learners still heavily rely on generic spell checkers for proofing their foreign language texts. Empirical evaluations of the efficacy of generic spell checkers are thus necessary.

THE STUDY

To evaluate a generic spell checker on nonnative misspellings and to investigate factors that may influence spell-checking results, we first classified a corpus of misspellings into several error categories and determined the frequency of each category. The corpus of misspellings was then used to determine the performance of the spell checker in treating these errors. The following subsections describe the methodology used in our study for data collection, error identification and classification, and examination of the spell checker's performance.

Data Collection

Participants

We collected misspellings from 48 students enrolled in one of the first two German language courses at a Western Canadian university (32 first-semester students, 16 second-semester students). According to a questionnaire distributed to all study participants at the beginning of the semester, the 30 female and 18 male students are all native English speakers with a mean age of 20.3 years. Ethics approval to participate in our study was obtained from our university's Office of Research Ethics and all participants agreed to participate at the beginning of their course.

Activity types

For data collection, we employed the *E-Tutor* (Heift & Nicholson, 2001; www.e-tutor.org), a parser-based CALL program that study participants used as part of their German courses. The *E-Tutor* logged each participant's misspellings. For each misspelling, the log provides information on student input and system response, student ID, exercise number, activity type, and access time.

Our corpus was collected from two different *E-Tutor* activities: translation and build-a-sentence. Translation and build-a-sentence exercises are commonly used in a CALL environment and thus representative of some of the learning activities that beginner and early intermediate students pursue. Moreover, the fairly controlled CALL activities facilitate the matching of misspellings with their respective target words, which is important in evaluating the spell checker's success in treating these misspellings.

The translation exercise (Figure 1) entails translating sentences from English into German. The build-a-sentence exercise (Figure 2) involves constructing sentences from a set of German words in their base forms. For example, for the sentence in (1a), students have to provide the simple past of the verb *haben*, determine the inflected indefinite pronoun *keine* (feminine, singular, accusative), and produce the correct word order, as given in (1)b.

(1)	a.	ich /	gestern /	kein- /	Zeit /	haben (simple past)
		I	yesterday	no (base form)	time	have
	b.	Ich	hatte	gestern	keine	Zeit.
		I	had	yesterday	no	time
		'I didn't have time yesterday.'				

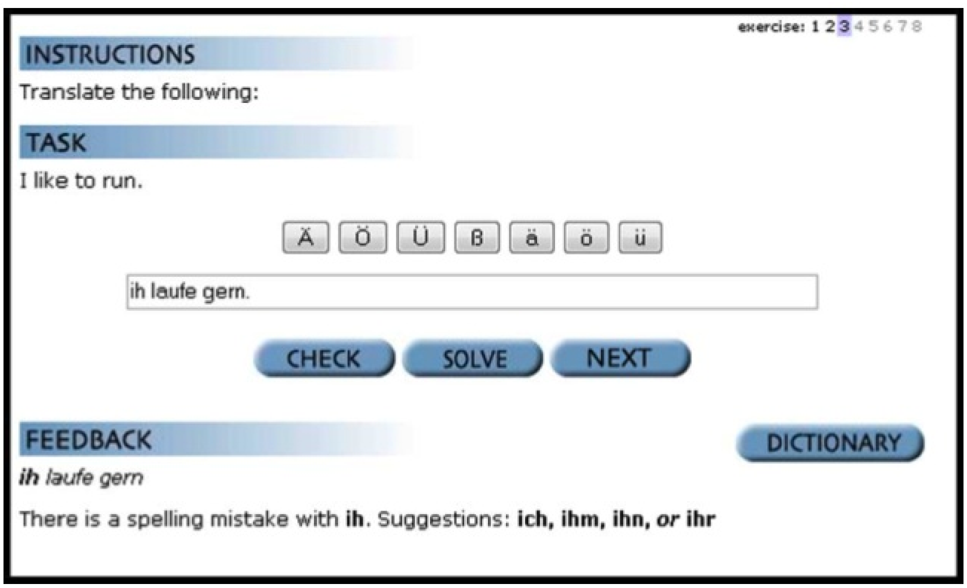

Figure 1. Translation activity in the *E-Tutor*

For both activity types, the user interface consists of an input field and buttons to (a) check the answer, (b) display the most common answers, and (c) advance to the following exercise. If students check their answer, a feedback message either tells them that the answer is correct or informs them of the type of mistake they made (e.g., a spelling mistake). An example of a spelling mistake is provided in Figures 1 and 2 (**ih/ich* 'I'). In the case of incorrect input, students may choose to revise and resubmit their answer, to look up possible answers, or to move on to the next exercise. Students also have access to an online bilingual dictionary.

As part of their coursework over 13 weeks, each participant completed five chapters of the *E-Tutor*. The first-semester students worked on chapters 1 to 5 of the *E-Tutor*, which contain 90 build-a-sentence and 34 translation exercises that present material from the first-semester syllabus (e.g., present tense, separable prefix verbs). The second-semester students completed chapters 6 to 10, which consist of 100 build-a-sentence and 25 translation exercises involving grammar pertaining to the second-semester curriculum (e.g., dative case, present perfect). Neither of the activity types is explicitly designed to elicit misspellings.

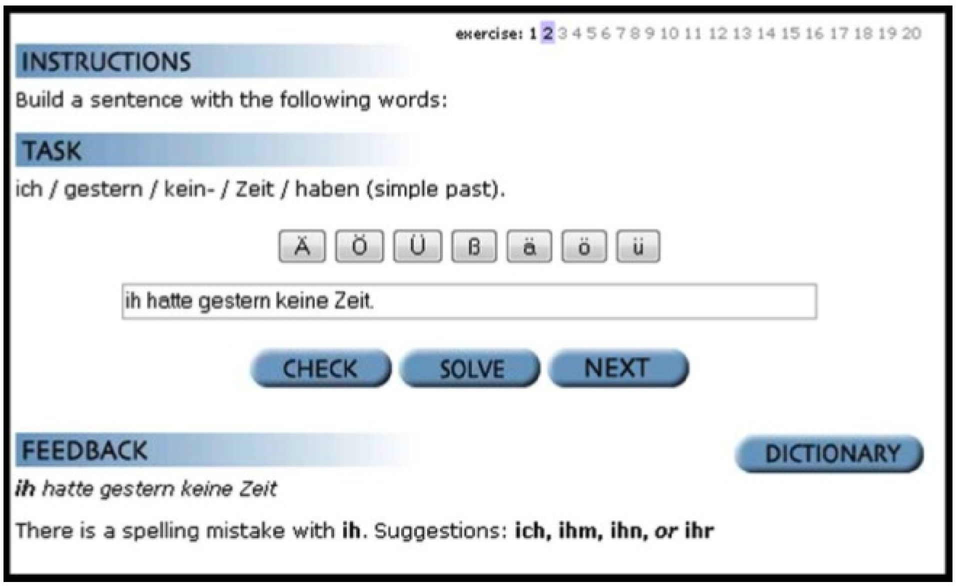

Figure 2. Build-a-sentence activity in the *E-Tutor*

Error Identification and Classification

To classify the misspellings that we collected from the two *E-Tutor* activities, we devised CLASSY, an error classification system that, in addition to shedding light on the different types of L2 misspellings, enabled us to investigate and pinpoint the types of L2 misspellings that are most problematic for a generic spell checker. Undoubtedly, a very fine-grained classification of misspellings is required to make improvements to L2 spell checking both from a pedagogical as well as computational point of view.

CLASSY

CLASSY consists of three error classification taxonomies, each comprising several error categories:

1. linguistic competence taxonomy: competence versus performance misspellings
2. linguistic subsystem taxonomy: lexical, morphological, or phonological misspellings
3. target modification taxonomy: single-edit versus multiple-edit misspellings

CLASSY primarily distinguishes between competence and performance misspellings according to the linguistic competence taxonomy. Competence misspellings are further classified as lexical, morphological, or phonological misspellings based on the linguistic subsystem taxonomy. In a last step, lexical, morphological, and phonological competence misspellings are categorized as single-edit or multiple-edit errors according to the target modification taxonomy. Performance misspellings, the second main error category, are only categorized based on the target modification taxonomy as single-edit or multiple-edit errors. The following subsections describe CLASSY in more detail.[4]

Linguistic competence taxonomy

The primary distinction in CLASSY is made between competence and performance misspellings according to the linguistic competence taxonomy. Competence errors involve misconceptions of target language forms and are due to a lack of linguistic knowledge on the writer's part. They are generally systematic, and/or non-self-corrigible and/or deliberate in that the erroneous form is assumed to be correct. Performance errors, on the other hand, are accidental, unsystematic and self-corrigible and can be attributed to factors like inattention or poor motor coordination. Corder (1975) refers to competence errors as *errors*, which are "typically produced by people who do not yet fully command some institutionalized language system (e.g. learners …)" (p. 204). He calls performance errors *mistakes*, stating they involve "failures to utilise a known system correctly" (p. 204).[5] James (1998) generally agrees with the competence/performance dichotomy but, at the same time, raises awareness that the notion of self-corrigibility can be problematic because even a competence error might be self-corrigible if explained to the learner (see pp. 76-86 for a nuanced discussion). These caveats notwithstanding, the distinction between competence and performance errors is useful in L2 misspelling studies because it captures the main difference between native and nonnative writers. While both native and nonnative writers produce performance errors, nonnative writers, due to a lack of L2 proficiency, also commit competence errors that adult native writers would not make (e.g., **goed/went*). The division between competence and performance errors is therefore central to this study.

Competence errors: linguistic subsystem classification

Competence-based misspellings are further classified into three linguistic subsystems: lexical, morphological, or phonological. Examples are provided in Table 2 and Appendix A.

Lexical misspellings are due to a lack of L2 lexical knowledge on the writer's part. The lack of lexical knowledge can be partial, as in the approximation of the correct spelling in, for instance, **Poskeutzah/Postleitzahl* 'postal code', or complete, as in the transfer of an English expression (e.g., **Suitcase/Koffer* 'suitcase').

Table 2. Linguistic Subsystem Taxonomy

Linguistic subsystem	Example
Lexical	**Suitcase/ Koffer* 'suitcase'
Morphological	**gebst/gibst* 'give'
Phonological	**biem/beim* 'at the'

Morphological misspellings involve difficulties with inflecting or deriving words. For example, strong verbs in German require a stem vowel change. In **gebst/gibst* 'give', the stem vowel change in the 2nd person singular of *geben* 'to give' is ignored.

Phonological misspellings contain cases where the actual or assumed phonology of a word affects its spelling. The German spelling of English native speakers can be influenced by both German and English phonology and spelling (James & Klein, 1994). For example, in **biem/beim* 'at the', the phoneme /aj/ is represented by the English grapheme *ie* instead of the correct *ei*.

Competence errors in the linguistic subsystems: Target modification classification

Competence misspellings of the three linguistic subsystem categories are further classified according to target modification. This taxonomy categorizes misspellings as single-edit or multiple-edit errors depending on their edit distance, that is, the number of character additions, omissions, substitutions, or transpositions needed to convert a misspelling into its target word.[6] Misspellings with an edit distance of one are single-edit errors (e.g., **gebst/gibst* 'give', substitution e/i). Multiple-edit misspellings have an edit distance of two or more (e.g., **gewascht/gewaschen* 'washed', edit distance 2: substitution t/e, addition -/n).[7]

Performance errors: target modification classification

Next to competence errors, performance errors are the other main category of nonnative misspellings. Performance errors are accidental by definition and as such do not involve misconceptions of linguistic subsystems. They are therefore subcategorized using only the target modification taxonomy described above. For example, **ds/das* 'the' is a single-edit performance misspelling, whereas **häaalich/hässlich* 'ugly' is a multiple-edit error.

Error classification procedures

Our corpus of misspellings, the *E-Tutor corpus*, contains 1027 types (unique misspellings), comprising 1808 tokens (non-unique misspellings). The procedure for collecting and classifying the misspellings from the *E-Tutor corpus* into CLASSY involved three main steps:

First, all misspellings, which were initially identified by the *E-Tutor*, were checked manually by the first author to ensure data accuracy. This was necessary because the *E-Tutor* occasionally overflagged words as misspellings (13 in total), which were spelled correctly but not contained in its dictionary (e.g., proper nouns). These submissions were removed from the corpus.

In the second step, the first author determined the writer’s intended target word for each misspelling. While this may present a challenge in more open activities (e.g., essays), our activity types were fairly constrained, which made it possible to deduce the target word from context. Yet, 11 misspellings for which the target word could not be determined unequivocally were excluded from the corpus. Additionally, 68 words were excluded because they contained more than one spelling error in the same word.

Thus, the *E-Tutor* initially flagged 1119 words, but as part of the first two steps of error collection and classification, 92 misspellings (13 + 11 + 68) were removed, yielding a working total of 1027 unique misspellings for the *E-Tutor corpus*. Note that the E-Tutor does not record a misspelling unless the CHECK button is clicked (see Figure 1). For this reason, our corpus reflects all misspellings that the students submitted to the E-Tutor for review, while the participants may have produced and self-corrected additional misspellings in the process of composing their response and prior to submitting it for analysis (Smith, 2008).

In the third step, each one of the 1027 misspellings was then categorized using CLASSY. Two coders, the first author and a native German speaker with a university degree in English and Linguistics, independently assigned the misspellings to their respective categories. The initial consensus in error category assignment between the coders was 94% (962/1027 misspellings). The remaining 6% were subsequently discussed by the coders to achieve a final consensus of 100%.[8]

Examination of the Spell Checker’s Performance

All misspellings from the *E-Tutor corpus* were pasted into a *Microsoft Word*® document (one misspelling per row) to assess the performance of the *Microsoft Word*® spell checker. The spell checker was then applied to each row at a time.[9]

RESULTS

Misspelling Distribution

To evaluate the performance of a generic spell checker on nonnative misspellings and to investigate what factors influence spell checking results, we first must determine the distribution of misspellings within

CLASSY. For this, we classified the 1027 unique misspellings of the *E-Tutor corpus* according to the procedures outlined above and determined their frequency.

Regarding linguistic competence, our results indicate that 72% (735) of the *E-Tutor*'s 1027 misspellings are competence errors and 28% (292) are performance errors. Within the competence category, the breakdown by linguistic subsystem in Table 3 shows that most competence misspellings are morphological (42%), followed by phonological (30%) and lexical (27%) errors.

Table 3. Distribution of Competence Misspellings across Linguistic Subsystem

Error category	Number of errors	Percentage of competence
Lexical	201	27%
Morphological	310	42%
Phonological	224	30%
Total	735	100%

Table 4 displays the target modification distribution of all misspellings. For instance, 95% (278) of the 292 performance misspellings are single-edit errors (i.e., their edit distance is one). The remaining 5% (14) are multiple-edit misspellings. In contrast, in the competence category, only 65% (477) are single-edit misspellings and 35% (258) are multiple-edit misspellings.

Table 4. Target Modification Distribution of All Misspellings

Linguistic competence	Target modification					
	Single-edit		Multiple-edit		Total	
Competence	477	65%	258	35%	735	100%
Linguistic subsystem						
Lexical	*112*	*56%*	*89*	*44%*	*201*	*100%*
Morphological	*155*	*50%*	*155*	*50%*	*310*	*100%*
Phonological	*210*	*94%*	*14*	*6%*	*224*	*100%*
Performance	278	95%	14	5%	292	100%
Total	755	74%	272	26%	1027	100%

Our results indicate that L2 misspellings differ substantially from typical L1 misspellings in that, for example, most L2 misspellings (72%) are *competence-based* whereas most L1 misspellings are *performance-based* (Damerau, 1964; Helfrich & Music, 2000; Pollock & Zamora, 1984). It is therefore necessary to evaluate the effectiveness of a generic spell checker on nonnative misspellings. Results of this evaluation are presented in the following subsection.

Spell Checker Evaluation

In pursuing our two research questions, we consider three possible spell checking outcomes in the evaluation of the generic spell checker in *Microsoft Word®*:

1. misspelling corrected,
2. misspelling uncorrected, and
3. misspelling undetected.[10]

The most successful result occurs when the spell checker detects a misspelling and provides the intended target word in its list of correction suggestions (*misspelling corrected*). In the second scenario, the spell checker detects the misspelling but does not suggest the target word as a correction (*misspelling uncorrected*). In this case, the spell checker either does not provide a list of suggestions, or the target word is not contained in the list. The least desirable outcome occurs when the spell checker fails to detect (and hence to correct) the misspelling (*misspelling undetected*).

Correction rate of L2 misspellings

Table 5 indicates that 62% (633/1027) of all L2 misspellings are corrected. The correction rates for competence misspellings (62%, 455/735) and performance misspellings (61%, 178/292) are almost equal. Within the competence category, Table 5 further illustrates that only 51% of the lexical and 47% of the morphological misspellings are corrected while 92% of phonological errors are corrected.

In addition to analyzing corrected and uncorrected misspellings, an evaluation of the spell checking outcome also comprises an investigation of undetected misspellings. Table 5 reveals that 6% of all misspellings and 5% of the competence misspellings are not at all detected.

Table 5. Spell Checking Results

Error category	Spell checking result							
	Corrected		Uncorrected		Undetected		Total	
Competence	455	62%	243	33%	37	5%	735	100%
Single-edit	442	93%	12	3%	23	5%	477	100%
Multiple-edit	13	5%	231	90%	14	5%	258	100%
Lexical	102	51%	80	40%	19	9%	201	100%
Single-edit	101	90%	2	2%	9	8%	112	100%
Multiple-edit	1	1%	78	88%	10	11%	89	100%
Morphological	147	47%	151	49%	12	4%	310	100%
Single-edit	146	94%	1	1%	8	5%	155	100%
Multiple-edit	1	1%	150	97%	4	3%	155	100%
Phonological	206	92%	12	5%	6	3%	224	100%
Single-edit	195	93%	9	4%	6	3%	210	100%
Multiple-edit	11	79%	3	21%	-	-	14	100%
Performance	178	61%	91	31%	23	8%	292	100%
Single-edit	178	64%	77	28%	23	8%	278	100%
Multiple-edit	-	-	14	100%	-	-	14	100%
Total	633	62%	334	33%	60	6%	1027	100%
Single-edit	620	82%	89	12%	46	6%	755	100%
Multiple-edit	13	5%	245	90%	14	5%	272	100%

Factors influencing the correction rate of L2 misspellings

Given the differences in the relative number of corrected misspellings in the different error categories (e.g., 47% morphological vs. 92% phonological), the question arises as to which factors influence the outcome of the spell checking process. The spell checking results for all of CLASSY's error categories, displayed in Table 5, show that generally, single-edit misspellings have a correction rate of 90% or higher whereas multiple-edit misspellings have a correction rate of 5% or lower. For instance, Table 5 reveals that 93% (442) of the single-edit competence misspellings are corrected but only 5% (13) of the multiple-edit ones. Two exceptions apply to these general findings:

1. Single-edit misspellings in the performance class have a low correction rate. Only 64% (178/278) of the single-edit performance errors are corrected.

2. Multiple-edit misspellings in the phonological category have a high correction rate. 79% (11/14) of multiple-edit misspellings in the phonological category are corrected.

As a corollary to these general findings, the spell checking success is different for each linguistic subsystem category because of different ratios of single-edit versus multiple-edit misspellings in each linguistic subsystem.

In the lexical category, 90% (101) of single-edit errors are corrected as opposed to 1% (1) of multiple-edit errors. The overall correction rate for lexical misspellings is 51% (102). Along similar lines, 94% (46) of single-edit morphological misspellings are corrected and 99% (154) of multiple-edit errors are not. The overall correction rate is 47% (147). In contrast, most phonological misspellings are corrected (92%, or 206) due to the high number of single-edit errors (94%, Table 4).

DISCUSSION

The results for our first research question indicate that the success rate of a generic spell checker is only 62% for the *E-Tutor* corpus. The correction rate is therefore considerably lower than the correction rates of above 90% that have been reported for L1 misspellings (Kukich, 1992). For performance misspellings, a low correction rate is less of a concern (as long as the misspellings are still detected) given that these errors are accidental and can easily be self-corrected by the writer. However, the low correction rate for competence misspellings (62%) is disconcerting because these errors are generally systematic, deliberate, and not self-corrigible, and, therefore, feedback that provides the language learner with the correct spelling of a target word is important. This is especially true considering the prevalence of competence misspellings in the *E-Tutor corpus* (72% of all misspellings).

While most competence misspellings that are not corrected by the spell checker are at least detected, there are also some competence misspellings (5%, 37/735) that remain undetected, that is, they are not flagged as misspellings by the spell checker. Closer examination of our data indicates that several reasons account for undetected misspellings (e.g., nondetection of some English words such as **France/Frankreich* 'France'). The most common reason, affecting 35% (13/37) of the misspellings, is the spell checker's treatment of German compounds. It is not feasible for a German spell checker to list all possible compounds in its dictionary because compounding is highly productive in German. Instead, spell checkers treat unknown word combinations composed of possible German words as correct spellings. For example, the spell checker did not detect the misspelling **Nachtmittag* (*Nacht* 'night' + *Mittag* 'noon') for *Nachmittag* 'afternoon' because, morphologically, it is a possible compound but not semantically.

With respect to our second research question, factors influencing correction success, our results indicate that the spell checking outcome in *Microsoft Word*® can largely be predicted based solely on the number of single-edit and multiple-edit misspellings contained in the corpus: single-edit misspellings are generally corrected, multiple-edit misspellings are generally not corrected. This is not surprising because, as stated earlier, generic spell checkers are designed for native speakers (Helfrich & Music, 2000), who mainly produce performance-related misspellings, the vast majority of which are single-edit errors (Pollock & Zamora, 1984; Damerau, 1964). These general findings, however, do not apply to single-edit performance misspellings and to multiple-edit competence misspellings in the phonological category.

The low correction rate for single-edit performance misspellings (64%, Table 5) is unexpected because, as stated above, *in theory*, generic spell checkers aim to correct single-edit misspellings of any kind. *In practice*, however, closer examination of the performance misspellings of the *E-Tutor corpus* reveals that the spell checker generally does not correct single-edit misspellings involving an incorrectly spelled first letter or non-letter characters. However, only 63% of the performance misspellings are single-edit errors with a correct first letter and no non-letter characters. In line with previous findings (Kukich, 1992), 92% of these misspellings are corrected.

Regarding the high correction rate for multiple-edit phonological misspellings, closer investigation of our data uncovers that the generic spell checker corrects confusions of the graphemes *ss* and *ß*, both of which represent the phoneme /s/ in German (e.g., **Fluß/Fluss* 'river'). These multiple-edit errors are presumably so frequent among native German speakers that their correction is built into the spell checker.

The strong relationship between edit distance and correction rate largely explains the correction rates in the corpus. Given that our L2 misspellings are mainly competence-related (72%) and as such often multiple-edit errors, the total correction rate (62%) and the correction rate for competence misspellings (62%) are low.

Within the subcategories of the competence category, that is, regarding lexical, morphological and phonological misspellings, we observe a very similar correlation. The correction rate for lexical misspellings is low (51%) because of the high number of multiple-edit errors in the lexical category (44%). Lexical misspellings are frequently multiple-edit errors because they are due to insufficient knowledge of entire words (as opposed to just morphemes or phonemes). The lexical misspelling **Poskeutzah/Postleitzahl* 'postal code' (edit distance: 4), for instance, demonstrates that lexical misspellings often deviate considerably from their target words, which makes it difficult for the spell checker to correct them.

Morphological misspellings also have a low correction rate (47%) because of numerous multiple-edit misspellings in the morphological category (50%). Morphological misspellings are frequently multiple-edit errors because morphemes are often composed of several letters. For example, *gehen* 'to go' has an irregular participle. A common mistake is to overgeneralize the present tense stem *geh* and generate **gegehen* instead of *gegangen* 'went', which leads to a multiple-edit error. Again, **gegehen* is so different from its target word *gegangen* (edit distance: 3) that the spell checker is unable to provide the correction.

In contrast, phonological misspellings have a high correction rate (92%) because of a high rate of single-edit errors (94%) and the fact that multiple-edit phonological misspellings are generally also corrected. Phonological misspellings are much more likely to result in single-edit errors than lexical or morphological misspellings because most phonemes in German are represented by graphemes consisting of one or two letters. For the single-edit misspelling **diser/dieser* 'this' (edit distance: 1), for instance, the spell checker successfully suggests the target word *dieser* as a correction.

Implications

Our study suggests that a large number of L2 misspellings are not adequately addressed by generic spell checkers because they are multiple-edit errors and differ from typical L1 misspellings. Given the low correction rate of 62%, both overall and for competence misspellings, this finding prompts several computational and pedagogical suggestions to enhance L2 spell checking. Along the lines of Tschichold’s (1999) strategies for improving L2 grammar checking, we suggest both computational and pedagogical strategies for improving L2 spelling, discussed next.

Computational suggestions

The results of our evaluation of a sophisticated, yet generic, spell checker demonstrate that spell checkers require additional algorithms that specifically target nonnative misspellings. It is not feasible to implement a separate algorithm for each and every misspelling. Instead, four equally important factors should guide computational efforts:

1. frequency: frequent errors have priority over infrequent errors
2. correction rate: errors with a low correction rate have priority over errors with an already high correction rate
3. predictability: predictable errors have priority over unpredictable errors

4. source: competence errors have priority over performance errors.

Accordingly, Table 6 presents recommendations for each of the three linguistic subsystem categories of competence errors.

Table 6. Recommendations for L2 Spell Checking

Error category	Frequency (in corpus)	Correction Rate (by spell checker)	Predictability	Recommendation
Lexical	high (27%)	low (51%)	low	not feasible to target
Morphological	high (42%)	low (47%)	high	target primarily
Phonological	high (30%)	high (92%)	high	need not be targeted primarily

Table 6 indicates that L2 spelling algorithms for German should be primarily directed at the correction of morphological misspellings because of their high frequency, low correction rate, and high predictability. Morphological misspellings are highly predictable due to the overgeneralization of inflectional paradigms (e.g., regularizing the irregular past tense *went* to **goed*). In the *E-Tutor corpus*, the correction rate for competence misspellings would improve from 62% to 82% (+ 151/735) if all uncorrected morphological misspellings were corrected. To correct morphological misspellings, spell checkers need a morphological analyser that recognizes erroneous but systematic misspellings (e.g., **goed*/*went*) and then generates the target form (e.g., *went*). In contrast to generic spell checkers, a spell checker with a morphological analyzer could provide learner feedback that addresses the morphological cause of the error instead of focusing on the nonexistent spelling.

Furthermore, Table 6 shows that lexical misspellings are too unpredictable to warrant computational efforts because a lack of L2 lexical knowledge can manifest itself in a number of distinct ways. For example, the student might add, delete, substitute and/or transpose several letters, with the exact nature of the transformation difficult to predict (e.g., compare two lexical misspellings for *Postleitzahl* 'postal code': *Poskeutzah*, *Postelizt*).

Finally, Table 6 shows that there is no pressing need to target phonological misspellings because of an already high correction rate.

Pedagogical suggestions

The second strategy for improving spell checking in CALL is to increase learners' ICT (Information and Communication Technology) literacy and decrease their dependence on the spell checker. In contrast to the computational suggestions, the pedagogical strategies do not require new spell checking algorithms. Instead, they can be more easily integrated into the foreign language classroom. Regarding CALL environments, three main points are worth considering:

First, more attention must be paid to ICT literacy training. Granger and Meunier (1994) suggested many years ago that L2 proofing tools should provide clear information on what they can and cannot do. Yet, even nowadays few computer programs train users on the general workings and limitations of the tool as opposed to providing mere operational guidance. Regarding spell checking, this training could include teaching spell checker limitations that are applicable to all writers (e.g., the target word may be absent from the spell checker's suggestion list), and issues particularly pertinent to L2 writers. For example, spell checkers should inform their users that they identify *all* nonexistent words even if the underlying cause is not spelling-related (as in morphological and lexical errors). They should also reveal that while they are able to *detect* these errors, their ability to *correct* morphological and lexical errors is low. Along those lines, learners should be aware that multiple-edit misspellings in general are not adequately corrected by generic spell checkers.

As a consequence, a second strategy is to offer language learners additional resources relevant to L2 writing to overcome some of the limitations of spell checkers (see also Tschichold, 1999), particularly the lack of correction of multiple-edit lexical and morphological misspellings. Given that only 51% of our lexical errors are corrected, students can be encouraged to consult dictionaries instead of solely relying on the spell checker. Furthermore, the low correction rate for morphological errors (47%) suggests that learners could benefit from access to morphological paradigms, for example, through dictionaries structured according to word formation rules (see, e.g., ten Hacken & Tschichold, 2001, and www.canoo.net). Also, both cross-linguistic and learner corpora could prove helpful to language learners. Ideally, these additional resources are made readily accessible to the learners by including them in course materials and websites. Again, students need to be trained in using these tools most effectively.

As a third point of improvement, the amount of misspellings students produce in the first place can be reduced. For example, we can teach learners more about L2-specific phoneme-grapheme correspondences (e.g., *ie* versus *ei* in German) and, more generally, orthography. The CALL program *eSpindle* (www.espindle.org, reviewed by Olmanson, 2007), for instance, helps learners of English practice spelling (see also Nicholas, Debski & Lagerberg, 2004). In addition, typical L2 competence misspellings can be discussed in the classroom. These strategies may also lead to more successful long-term self-monitoring of students' writing (see Burston, 2001).

CONCLUSION

This article presented an evaluation of a generic spell checker on a corpus of nonnative misspellings by German learners. Overall, only 62% of the 1027 misspellings were successfully corrected. More importantly, only 62% of the 735 nonnative competence errors were corrected, confirming previous results of a small-scale study by Rimrott and Heift (2005). Moreover, the current study also found that the ratio of single-edit to multiple-edit misspellings in a given misspelling corpus is a strong predictor of spell checking success in that, generally, single-edit misspellings are corrected successfully while multiple-edit ones are not. Other factors such as the linguistic subsystem category of competence misspellings appear to be only influential inasmuch as they bear on the ratio of single-edit versus multiple-edit misspellings.[11]

Our study demonstrates that while the generic spell checker serves its primary purpose of correcting single-edit misspellings, most misspellings in nonnative writing are competence-based and thus frequently multiple-edit errors. The low correction success for competence misspellings suggests both a need to design spell checkers that specifically target L2 misspellings and a need to increase language learners' ICT literacy by also making students less reliant on spell checkers in L2 writing.

While our findings are revealing, there is room for future research. For example, although build-a-sentence and translation activities are common in CALL settings, language learners engage in a variety of other activities that might influence L2 misspellings in different ways. Kukich (1992) notes that "spelling error patterns vary greatly depending on the application task" and therefore, "care must be taken not to overgeneralize findings when discussing spelling error patterns" (p. 387). Accordingly, future studies might investigate how the results obtained here compare to other activity types such as free composition, not least because a free composition activity reflects what learners ultimately aspire to in L2 writing. Our study results, however, show that the main predictor of spell checking success is the ratio of single-edit versus multiple-edit errors because the algorithms of generic spell checkers mainly target single-edit misspellings, that is, the misspelling category most frequent in native speaker writing. This part of our findings is thus independent of activity and learner variables. Yet, what remains of interest and calls for further investigation is whether the distribution of single-edit versus multiple-edit errors for nonnative misspellings is comparable across activity types.

NOTES

1. For a more detailed discussion on spell checking algorithms and problems, see Kukich (1992).

2. For brevity, a discussion of L2 grammar checkers is omitted in this article. Instead, the reader is referred to Heift and Schulze (2007), who provide a comprehensive overview of existing grammar checkers for CALL.

3. Note that studies without a computational component (e.g., Rogers, 1984), classify errors such as, for example, **goed/went* as morphological errors because, from a Second Language Acquisition (SLA) point of view, the error is caused by overgeneralization of a morphological rule and not a misconception of its spelling. However, we follow the tradition in computational linguistics for the reasons stated.

4. CLASSY employs a fourth taxonomy, language influence, which we omit for brevity (see Rimrott, 2005, for details). CLASSY is similar to general error classification models such as the one described by van Els, Bongaerts, Extra, van Os and Janssen-van Dieten (1984). Their model uses linguistic competence as a primary distinction and further subdivides competence errors according to language influence and linguistic subsystem (see also James, 1998). Other nonnative misspelling studies have also served as a basis for CLASSY (e.g., linguistic competence: Staczek, 1982; Ibrahim, 1978; linguistic subsystem: James, Scholfield, Garrett & Griffiths, 1993; Snyder, 1995; target modification: Cook, 1997). Finally, instead of classifying the misspellings according to their *error category* as implemented in CLASSY, Brown (1970) classifies them according to spelling regularity and frequency of the *intended target word*.

5. Here, we use the terms *error* and *mistake* interchangeably, but the distinction between *competence* and *performance* is made consistently.

6. Note that edit distance is a computational algorithm that determines the minimum number of changes needed to transform a misspelling into its target word. The algorithm has no knowledge of the target language and thus entirely ignores possible reasons for why or how a misspelling was produced.

7. In this study, the target modification taxonomy is applied to misspellings in which *only one spelling error* occurs although it might require *more than one letter* to convert the misspelling into the target word. The taxonomy does not refer to misspellings that contain more than one distinct spelling error. Initially, our corpus included 68 misspellings that contained several distinct spelling errors. For instance, the misspelling **gewasht*/*gewaschen* 'washed' contains two spelling errors: a morphological error (the past participle is inflected with the *-t* suffix for weak verbs instead of the correct *-en* suffix for strong verbs) and a phonological error (the phoneme /ʃ/ is written with the English grapheme *sh* instead of the correct German *sch*). For clarity, the 68 misspellings that contained several distinct spelling errors were excluded from our study (for a discussion of such misspellings, see Rimrott & Heift, 2005).

8. In addition to more general classification procedures, a good classification system also requires that errors be categorized in a reliable fashion. For this, classification guidelines that "allow a reasonable, consistent, and meaningful analysis of the data" are essential because "like all other error data, … [misspellings] may permit multiple, even contradictory analyses" (Snyder, 1995, p. 103). CLASSY employed four diagnostics that assisted in assigning misspellings to their error categories: frequency, edit distance, systematicity, and previous research findings (see Appendix B). Overall, the four diagnostics allowed for a consistent error classification. However, as with any error classification system, and in the absence of think-aloud protocols or retrospective interviews, some degree of ambiguity cannot be ruled out completely given the many variables involved (e.g., learners, activity types).

9. The 2003 PC version of the spell checker was set to standard German and the default settings were used.

10. A final possibility is that the spell checker misidentifies an existing spelling (e.g., a proper name) as a misspelling. However, this possibility is excluded in this study because the misspellings in the *E-Tutor corpus* are all non-existent words.

11. However, factors independent of the error category of a misspelling, for instance, the learners' proficiency level or activity type, might also influence the spell-checking outcome (for an investigation of this, see Rimrott, 2005).

APPENDICES

Appendix A: Examples of Misspellings

Misspelling	**Target Word**	**Gloss**	**Category**	**Position**	**Length**	**Students**	**Frequency**
Austria	Österreich	Austria	IL2	0	0	11	16
Schestwe	Schwester	sister	IL2	0	1	1	1
Postelizt	Postleitzahl	postal code	IL2	0	0	1	1
Suitcase	Koffer	suitcase	IL2	-1	-1	1	1
aufgesteht	aufgestanden	got up	IM2	0	0	8	12
fahrst	fährst	drive	IM1	5	6	5	6
gefliegt	geflogen	flown	IM2	0	1	9	10
heißst	heißt	are called	IM1	5	5	9	21
Fluß	Fluss	river	IP2	1	6	2	3
irh	ihr	her	IP1	2	3	5	6
seiben	sieben	seven	IP1	9	9	3	6
Geshäfte	Geschäfte	stores	IP1	1	1	2	2
Bbend	Abend	evening	II1	0	2	1	1
3gesund	gesund	healthy	II1	-1	-1	1	1
ht	hat	has	II1	1	1	1	1
Won\her	Woher	where from	II2	0	0	1	1

Note. Category = Error category. I = Competence, II = Performance, L = Lexical, P = Phonological, M = Morphological, 1 = Single-edit, 2 = Multiple-edit. Examples: IM1 = Competence, Morphological, Single-edit, II2 = Performance, Multiple-edit. Position = Position of target word in spell checker's correction list. 1 or higher = position of target word in list, 0 = misspelling uncorrected, -1 = misspelling undetected. Length = Length of spell checker's correction list. 0 or higher = list length (in number of words), -1 = misspelling undetected (i.e., no list). Students = Number of distinct students that produced the misspelling, Frequency = Number of times the misspelling was produced.

Appendix B: Classification Guidelines

To assign misspellings to the main error categories of competence and performance, the following four diagnostics were applied to the *E-Tutor corpus*: frequency, edit distance, systematicity, previous research findings. Note that a misspelling was assigned to the competence category if any one of the four diagnostics supported this assignment; otherwise it was classified as performance-related. For consistency, identical misspellings were always classified as belonging to the same category.

Frequency

Performance errors are accidental by definition and, for this reason, *frequently* occurring identical errors – generally, errors that occurred more than five times – were classified as competence-related. For example, the grapheme *ei* was misspelled as *ie* over 70 times (e.g., in **biem/beim* 'at the') and misspellings containing this error were thus classified as competence-related. On the other hand, the grapheme *eu* was misspelled as *ue* only three times (e.g., in **Duetsch/Deutsch* 'German') and, therefore, these misspellings were classified as performance-related.

Edit distance

Regarding edit distance, L1 studies indicate that performance misspellings usually involve an edit distance of one (Damerau, 1964; Pollock & Zamora, 1984). To give the benefit of the doubt to performance, we classified misspellings with an edit distance of at least three as competence-related.

Systematicity

Misspellings involving systematic deviations from the target word were classified as competence-related. For instance, German verbs normally add *st* to their stem in the second person singular (e.g., *geh*-st '(you) go'), but verb stems ending in *d* or *t* additionally require *e* epenthesis (e.g., *find-e-st* '(you) find'). When learners systematically omit the *e* (e.g., **findst*/*findest*), the error is classified as competence-related.

Previous research findings

Errors that have been identified as typical of German learners were categorized as competence-related. For example, Rogers (1984) recognizes the confusion of *s*, *ss*, and *ß* as characteristic of L2 German.

ACKNOWLEDGMENTS

This research was supported by Social Sciences and Humanities Research Council (SSHRC), Canada, grant 632209, awarded to the second author. The authors would like to thank Kathrin Mink for helping to classify the misspellings and the participants for taking part in this research project. Thank you also to the anonymous reviewers and the LL&T editors for their valuable and insightful comments on earlier drafts of the manuscript.

ABOUT THE AUTHORS

Anne Rimrott is a Ph.D. student in the Department of Linguistics at Simon Fraser University, Canada. Within Computer-Assisted Language Learning, her research interests include vocabulary learning, student modeling and spell checking.

E-mail: arimrott@sfu.ca

Trude Heift is associate professor in the Department of Linguistics at Simon Fraser University. Her work bridges applied and computational linguistics by studying different aspects of learner-computer interaction in adaptive parser-based CALL systems.

E-mail: heift@sfu.ca

REFERENCES

Aspell 0.60 [Computer software]. Accessed August 22, 2008, http://aspell.net/

Bon Patron [Computer software]. Accessed August 22, 2008, http://bonpatron.com/

Bos, E. (1994). Error diagnosis in a tutoring system for the conjugation and spelling of Dutch verbs. *Computers in Human Behavior, 10*(1), 33-49.

Brown, H. D. (1970). Categories of spelling difficulty in speakers of English as a first and second language. *Journal of Verbal Learning and Verbal Behavior, 9*, 232-236.

Burston, J. (1998). Antidote 98 [review]. *CALICO Journal*, *16*(2), 197-212.

Burston, J. (2001). Exploiting the potential of a computer-based grammar checker in conjunction with self-monitoring strategies with advanced level students of French. *CALICO Journal*, *18*(3), 499-515.

Canoo-Net [Computer software]. Accessed August 22, 2008, http://www.canoo.net/

Cook, V. J. (1997). L2 users and English spelling. *Journal of Multilingual and Multicultural Development, 18*(6), 474-488.

Corder, S. P. (1975). Error analysis, interlanguage and second language acquisition (survey article). *Language Teaching and Linguistics. Abstracts, 8*(4), 201-218.

Cowan, R., Choi, H. E., & Kim, D. H. (2003). Four questions for error diagnosis and correction in CALL. *CALICO Journal*, *20*(3), 451-463.

Dagneaux, E., Denness, S., & Granger, S. (1998). Computer-aided error analysis. *System, 26*(2), 163-174.

Damerau, F. J. (1964). A technique for computer detection and correction of spelling errors. *Communications of the ACM, 7*(3), 171-176.

de Haan, A., & Oppenhuizen, T. (1994). Speller: A reflexive ITS to support the learning of second language spelling. *Computers in Human Behavior, 10*(1), 21-31.

eSpindle [Computer software]. Accessed August 22, 2008, http://www.espindle.org/

E-Tutor [Computer software]. Accessed August 22, 2008, http://www.e-tutor.org/

Fallman, D. (2002). *The Penguin: Using the web as a database for descriptive and dynamic grammar and spell checking.* Paper presented at the CHI 2002, Conference on Human Factors in Computing Systems, Minneapolis, MN, April 20-25.

Granger, S., & Meunier, F. (1994). Towards a grammar checker for learners of English. In U. Fries, G. Tottie & P. Schneider (Eds.), *Creating and using English language corpora: Papers from the fourteenth international conference on English language research on computerized corpora, Zürich 1993* (pp. 79-91). Amsterdam: Rodopi.

Heift, T., & Nicholson, D. (2001). Web delivery of adaptive and interactive language tutoring. *International Journal of Artificial Intelligence in Education*, *12*(4), 310-325.

Heift, T., & Rimrott, A. (2008). Learner responses to corrective feedback for spelling errors in CALL. *System, 36*(2), 196-213.

Heift, T., & Schulze, M. (2007). *Errors and intelligence in Computer-Assisted Language Learning: Parsers and pedagogues.* New York: Routledge.

Helfrich, A., & Music, B. (2000). Design and evaluation of grammar checkers in multiple languages. In *Proceedings of Coling 2000: The 18th international conference on computational linguistics* (Vol. 2, pp. 1036-1040): Association for Computational Linguistics.

Holmes, G., & de Moras, N. (1997). A French language grammar analyzer: What use for Anglophone students? In P. Liddell, M. Ledgerwood & A. Iwasaki (Eds.), *FLEAT III: Foreign language education and technology - proceedings of the third conference* (pp. 91-106). Victoria, British Columbia, Canada: University of Victoria.

Hovermale, D. J. (2008). *SCALE: Spelling correction adapted for learners of English.* Poster presentation at ICALL Special Interest Group pre-conference workshop, CALICO conference, March 18 – 22.

Ibrahim, M. H. (1978). Patterns in spelling errors. *English Language Teaching, 32*(3), 207-212.

James, C. (1998). *Errors in language learning and use: Exploring error analysis*. London: Longman.

James, C., & Klein, K. (1994). Foreign language learners' spelling and proofreading strategies. *Papers and Studies in Contrastive Linguistics, 29*, 31-46.

James, C., Scholfield, P., Garrett, P., & Griffiths, Y. (1993). Welsh bilinguals' English spelling: An error analysis. *Journal of Multilingual and Multicultural Development, 14*(4), 287-306.

Kese, R., Dudda, F., Heyer, G., & Kugler, M. (1992). Extended spelling correction for German. In *Proceedings of the third conference on applied natural language processing (ANLP), Trento, Italy* (pp. 126-132): Association for Computational Linguistics.

Kukich, K. (1992). Techniques for automatically correcting words in text. *ACM Computing Surveys, 24*(4), 377-439.

L'haire, S. (2007). Fipsortho: A spell checker for learners of French. *ReCALL*, *19*(3), 137-161.

Murphy-Judy, K. (2003). Sans-faute writing environment [review]. *CALICO Journal*, *21*(1), 209-220.

Mydlarski, D. (1999). Cortext [review]. *CALICO Journal*, *16*(4), 584-595.

Ndiaye, M., & Vandeventer Faltin, A. (2003). A spell checker tailored to language learners. *Computer Assisted Language Learning*, *16*(2-3), 213-232.

Nicholas, N., Debski, R., & Lagerberg, R. (2004). Skryba: An online orthography teaching tool for learners from bilingual backgrounds. *Computer Assisted Language Learning, 17*(3-4), 441-458.

Olmanson, J. (2007). Review of eSpindle vocabulary & spelling program online. *Language Learning & Technology, 11*(3), 18-28.

Pollock, J. J., & Zamora, A. (1984). Automatic spelling correction in scientific and scholarly text. *Communications of the ACM, 27*(4), 358-368.

Rimrott, A. (2005). *Spell checking in Computer-Assisted Language Learning: A study of misspellings by nonnative writers of German.* Master's thesis, Simon Fraser University, Canada.

Rimrott, A., & Heift, T. (2005). Language learners and generic spell checkers in CALL. *CALICO Journal, 23*(1), 17-48.

Rogers, M. (1984). On major types of written error in advanced students of German. *International Review of Applied Linguistics, 22*(1), 1-39.

Smith, B. (2008). Methodological hurdles in capturing CMC data: The case of the missing self-repair. *Language Learning & Technology, 12*(1), 85-103.

Snyder, W. E. (1995). *Cognitive strategies in second language lexical processing: Evidence from English speakers' spelling errors in Spanish.* Unpublished doctoral dissertation, Northwestern University, Evanston, IL.

Spellcheckplus [Computer software]. Accessed August 22, 2008, http://spellcheckplus.com/

Staczek, J. J. (1982). Expanded subcategorization of Spanish-English bilingual spelling strategies. In J. A. Fishman & G. D. Keller (Eds.), *Bilingual education for Hispanic students in the United States* (pp. 139-150). New York: Teachers College Press.

ten Hacken, P., & Tschichold, C. (2001). Word manager and CALL: Structured access to the lexicon as a tool for enriching learners' vocabulary. *ReCALL Journal, 13*(1), 121-131.

Tschichold, C. (1999). Grammar checking for CALL: Strategies for improving foreign language grammar checkers. In K. Cameron (Ed.), *Computer assisted language learning (CALL): Media, design and applications* (pp. 203-221). Lisse, Holland: Swets & Zeitlinger.

van Els, T., Bongaerts, T., Extra, G., van Os, C., & Janssen-van Dieten, A. M. (1984). *Applied linguistics and the learning and teaching of foreign languages* (R. R. van Oirsouw, Trans.). London: Edward Arnold.